5.4 A 2.6GHz Dual-Core 64b×86 Microprocessor with DDR2 Memory Support

Michael Golden, Srikanth Arekapudi, Greg Dabney, Mike Haertel, Stephen Hale, Lowell Herlinger, Yongg Kim, Kevin McGrath, Vasant Palisetti, Monica Singh

Advanced Micro Devices, Sunnyvale, CA

A microprocessor featuring 2 Hammer [1] cores and an on-chip DDR2 memory controller, and implementing Pacifica architectural support for virtualization has been fabricated in a 90nm triple-V_t partially-depleted SOI process with 9 layers of copper interconnect [2]. The chip achieves a clock frequency of 2.6GHz at 1.35V in a 95W power envelope. Figure 5.4.1 summarizes processor characteristics. 2 of the 9 metal layers are dedicated to a power and a ground plane, which provide low-impedance power rails and an inductive current-return path. The top level of metal is reserved for C4 landing pads, the clock grid, and a ground grid, leaving 6 layers for general routing.

Chip architecture and design methodology supported dual-core capability at the time of the initial single-core 0.13µm implementation [3]. Die area and power considerations make dual-core implementation viable with the 90nm process generation [4]. Most processor building blocks remain unchanged between single-core and dual-core versions of the chip. No changes are required to the core itself. I/O pads must be placed and routed differently. Logic area in the memory controller grows by 8% for dual-core support. Built-in redundancy, floorplanning of internal blocks to minimize timing and routing issues in both dual- and single-core modes, and routing spare wires in the single-core design to be connected to signals in dual-core design help converge the dual-core design at faster rate with fewer resources. SoC integration techniques allow rapid tape out of chips with single or multiple cores and different cache sizes and memory interface widths.

3 identical PLLs generate on-chip clocks from a 200MHz system clock input. 2 PLLs provide clocks for 3 Hyper Transport links, and the third provides a clock for the memory controller and both cores. The clock grid over the memory controller and the 2 cores can be separately enabled. A balanced H-tree drives the clock signal from the PLL to final clock buffers, which drive a uniform metal grid from the edges of 9 "panes," 4 over each core and 1 over the memory controller. Core clock panes are noted on the left-hand core in Fig. 5.4.2.

Extensive use of fine-grained clock gating reduces the load on the clock grid and reduces power consumption. The clock grids over the 2 cores can be separately enabled. In low-power operating modes, the clock grids over the CPU cores are disabled and the clock grid over the memory controller runs at $1/256^{th}$ the frequency of the system clock. The core clock is not driven over the L2 cache array. The L2 cache control logic receives the clock at the top edge of the array and drives timing control signals into the cache as needed.

The grid provides a low-resistance path to all clock receivers so clock drivers do not have to be tuned based on loading at the end of the design cycle. Final timing closure includes skew numbers calculated from a simulation of the clock net, including extracted parasitic resistance and capacitance, shown in Fig. 5.4.3. Worst-case clock skew is 21ps, and occurs along the border of the L2 cache, near the memory controller. Because this late skew causes no timing violations, the design tolerates it. There is considerably less skew in the rest of the core.

To transfer data to and from the DDR1 interface, previous implementations generate a low-skew clock distributed over portions of the I/O ring and the L2 cache. Data is sent from the memory controller to the pads synchronously via repeaters and latches placed in the L2 cache. This does not scale well to the higher clock speeds of DDR2 and the larger die area of a dual-core chip. Instead, data is delivered to the DDR2 interface from the memory controller via source-synchronous transfer in a routing channel. The removal of the area reserved for flops and repeaters allows a 6.5% reduction in L2 cache area.

Data and strobe signals written out to the memory subsystem have a timing relationship controlled by 2× MEMCLK, shown in Fig. 5.4.4. Careful transistor sizing and layout minimizes channel variance, duty-cycle uncertainty, clock jitter, and IR-drop difference between data and clock. Nonetheless, skew accumulates as data and clock travel across the chip. Periodic insertion of a clock retimer prevents setup and hold violations at the receiver.

Data outputs switch on the rising edge of 2× MEMCLK and strobe outputs switch on the falling edge. To remove the necessity of controlling the duty cycle of 2× MEMCLK, a DLL triggered off of the rising edge of 2× MEMCLK controls the timing relationship. The DLL sets the center point of a programmable delay element so that strobe and data are 90° out of phase. A separate DLL centers each clock signal being driven on or off the chip. The delay element can be fine tuned via BIOS programming either statically or dynamically based on training.

The chip implements the Pacifica architecture for hardware support of virtualization. Virtualization allows guest operating systems to operate under the supervision of a host hypervisor [5]. The guest operating system may or may not know of the existence of the hypervisor, and the hypervisor itself may coexist with an operating system or exist on its own as a thin layer. The hypervisor intercepts memory activity generated by guests, including DMA transactions, and processes interrupts. Changes from previous core versions to implement the Pacifica architecture are minimal. All new instructions are implemented in microcode, requiring a small increase in on-chip ROM size. Lengthening tags in the translation lookaside buffers by 6b for address space IDs, allows page mappings from multiple guests and the hypervisor to co-exist. Minor changes to the logic permit interrupt handling and DMA transactions.

The shmoo plot in Fig. 5.4.5 taken from a sample part demonstrates that the design has 7% frequency margin and 10% voltage margin at its operating point. At this operating point, the part consumes 95W. Large servers pack many processors into a small volume, and removing heat from the system becomes challenging. Reducing power dissipation by lowering voltage, without adversely affecting performance, ameliorates this problem. Figure 5.4.6 shows static leakage versus frequency for a population of parts tested at various supply voltages. Assuming that dynamic current is proportional to $C_{ac} \cdot V_{dd}^2 \cdot F$, reducing voltage from 1.35V to 1.1V achieves a three-fold reduction in static leakage and a 47% reduction in dynamic leakage at a cost of 20% in frequency.

References:

[1] C. Keltcher, "The AMD Hammer Processor Core," *Hot Chips 14*, Aug., 2002.

[2] D. Greenlaw, et al., "Taking SOI Substrates and Low-k Dielectrics into High-Volume Microprocessor Production," *IEDM Technical Digest*, pp. 11.1.1-11.1.4, Dec., 2003.

[3] C. Keltcher, et al., "The AMD Opteron™ Processor for Multiprocessor Servers," *Micro*, vol. 23, no. 2, pp. 66-76, Mar.-Apr., 2003.

[4] M. Evers, "Low Power AMD Athlon™ 64 and Opteron™ Processors," *Hot Chips 16*, Aug., 2004.

[5] P. Barham, et al., "Xen and the Art of Virtualization," *SOSP'03*, pp. 164-177, Oct., 2003.

 1-4244-0079-1/06/$20.00

	Model 854	Model 875	This work	This work
Process technology	90nm triple-Vt, partially-depleted SOI. 9 layer Cu metallization. Dual gate-oxide thickness.			
CPU Cores	1	2	1	2
Die area	106mm^2	194mm^2	126mm^2	220mm^2
L2 cache area	41.4mm^2	82.8mm^2	38.7mm^2	77.4mm^2
Transistor count L2 array L1 array	120 x 10^6 134 x 10^6 13 x 10^6	233 x 10^6 134 x 10^6 13 x 10^6	129 x 10^6 134 x 10^6 13 x 10^6	243 x 10^6 134 x 10^6 13 x 10^6
L1 instruction cache	64kB per core, parity protected			
L1 data cache	64kB per core, ECC protected			
L2 cache	1MB per core, ECC protected			
Memory interface	128b DDR1-400 6.4GB/s		128b DDR2-800 12.8GB/s	
Clock frequency	2.8GHz 1.35V 92.6W	2.2GHz 1.35V 95W	Data not available	2.6GHz 1.35V 95W

Figure 5.4.1: Comparison to previous AMD Opteron™ processors.

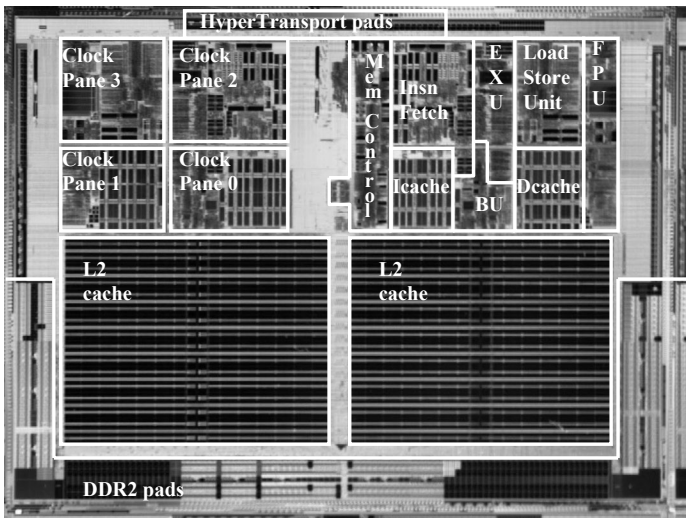

Figure 5.4.2: Annotated die micrograph.

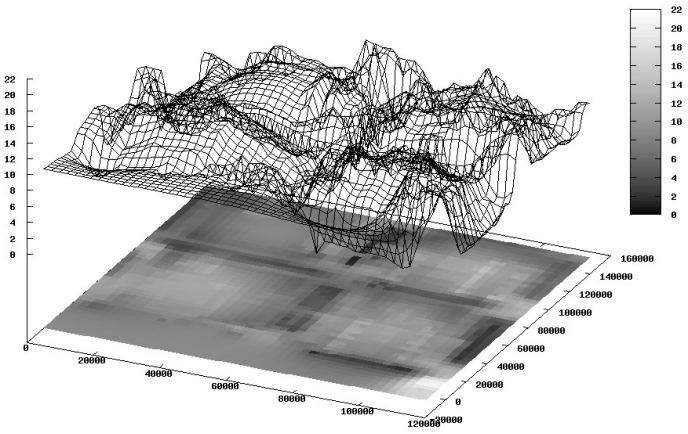

Figure 5.4.3: Clock skew map.

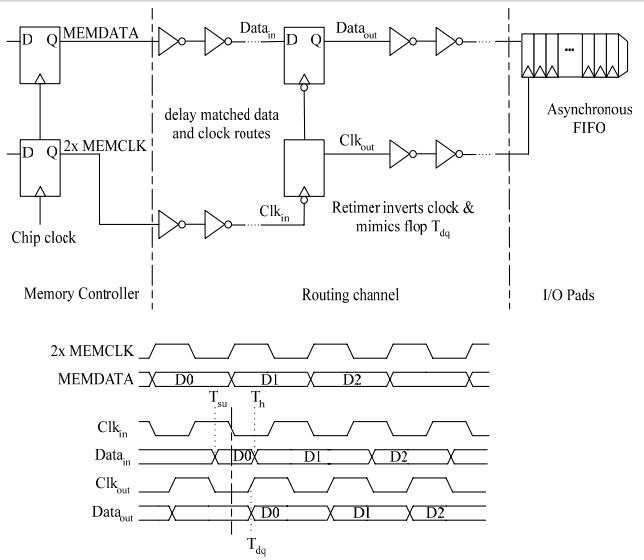

Figure 5.4.4: DDR2 retiming.

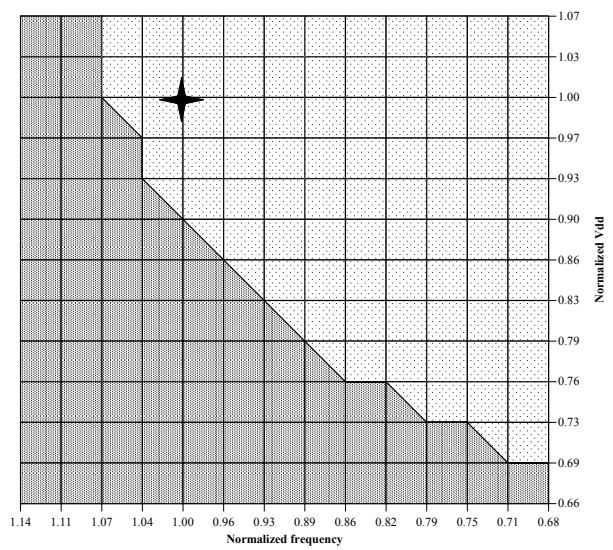

Figure 5.4.5: Frequency versus voltage shmoo.

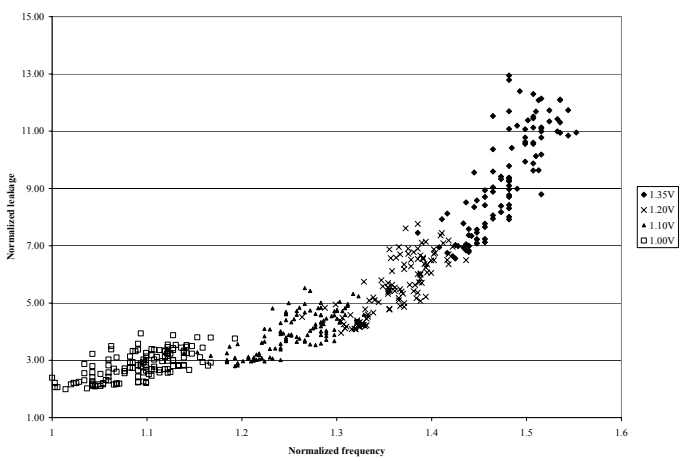

Figure 5.4.6: Static leakage versus frequency.